

*Gender bias in machine translation: an analysis of Google Translate in English and Spanish*

Maria Lopez Medel

---

More than seventy years after Warren Weaver's prediction, automatic translation has become a part of our daily lives, both personally and professionally, even for language specialists. Translation engines are used by millions of people every day, delivering the best results between grammatically-similar languages with larger multilingual corpora available, like the English-Spanish pair. Big translation departments also rely heavily on machine translation based on custom memories and databases, and translation processes have dramatically changed.

For non-professional users, a service like Google Translate offers draft-quality texts that are free, almost unlimited and faster than human translations. But automation comes at a price, specially with regard to unwanted bias, gender-wise and other.

Although Google has attempted on several occasions to remove gender bias from its free online translation service, it still tends to exhibit predominantly masculine options and shows a tendency towards perpetuating or exaggerating sexist stereotypes, which adds to other flaws like a failure to notice text formality, typos and nuances.

The technology on which its machine translation system is based, that feeds from statistics and large multilingual corpora such as the UN and the European Parliament, is supposed to reflect the gender divide in society, including men's and women's professional quotas, but the results can show even greater inequalities than in real life.

In 2018 Google introduced a development that showed gender-specific translations in a selection of languages for single words or even short phrases and sentences for fewer languages, like Spanish, but it was not devoid of a masculine default.

To counteract it, the company launched an update in 2020 based on its neural translation technology that generates a default translation, rewrites results that are gendered and checks

for accuracy. Now, translations between different-gender languages like Spanish and English should offer the feminine and masculine versions, but this is not always the case.

To analyse the gender bias of Google Translate between English and Spanish, we entered for translation a number of gender-invariable Spanish nouns whose referent's gender was unknown due to the omission of pronouns and other particles. In several cases, an invariable (genderless) adjective was added to the search in order to test for changes.

The sentences were so short that on occasions the pronoun chosen by the engine was neutral in English (*Es demandante>It is demanding*, instead of *He or she is a plaintiff*) and they were consequently discarded from the study. Since gender results varied with punctuation, all sentences were capitalised and a full stop added (*es colega>is a colleague*, but *Es colega.>He's a colleague.*)

Out of the 74 Spanish invariable nouns entered in the “*Es + noun*” format (*He or she is...*), only 4 adopted the feminine form in the English translation (*dressmaker, feminist, transgender and transexual*) while the rest were invariably masculine, unlike with longer and shorter sentences, which included both genders. Some gender attributions were unexpected (both *suffragette* and *beautician* were men) and there was one typographical error (*antenist\**). Another unusual mistake was *Es agente>Those people* (for *esa gente*), where the space had been wrongly moved.

<i>Sentence construction</i>	<i>Personal nouns</i>	<i>Gender</i>
<i>He is a/an...</i>	<i>beautician, colleague, contractor, criminal, dentist, dissident, donor, flute player, fundamentalist, journalist, lawyer, leader, linguist, listener, machinist, manager, Maoist, Marxist, member, Methodist, migrant, naturist, participant, pensioner, pianist, reformist, resident, ruler, shareholder, speaker, spokesperson, spouse, student, suffragette, survivor, taxi driver, teacher, teenager, terrorist, trade unionist, veteran, violinist, visitor, voter, American, activist, analyst, antenist*, athlete, economist, electrician, emigrant, immigrant, interpreter, Mapuche, militant, Nicaraguan, Professional, Roma, owner, protagonist, cop, guard</i>	<i>Masculine</i>
<i>She is a...</i>	<i>dressmaker, feminist, transgender, transexual</i>	<i>Feminine</i>

TABLE 1. Gender variations of “*Es + noun*” sentences translated from Spanish

A second search focused on gendered-career nouns whose translation changed while the word was being written. Either the translation was masculine and/or the engine suggested its completion in masculine form (*Es ingenier...>He is an engineer. Es medic...>He is a doctor*). Masculine seemed to be the default gender regardless of job stereotypes, men and women's quotas or alphabetical order (in Spanish all these words end in *a* for feminine and *o*

for masculine).

Spanish	Suggestion	Gender
<i>Es <del>camioner</del>...</i>	<i>He is a truck driver.</i>	<i>Masculine</i>
<i>Es secretari...</i>	<i>He is a secretary</i>	<i>Masculine</i>
<i>Es enfermer...</i>	<i>He is a nurse.</i>	<i>Masculine</i>
<i>Es ingenier...</i>	<i>He is an engineer.</i>	<i>Masculine</i>
<i>Es médic...</i>	<i>He is a doctor.</i>	<i>Masculine</i>
<i>Es <del>abogad</del>...</i>	<i>He is a lawyer.</i>	<i>Masculine</i>

TABLE 2. Gender suggestions to unfinished personal noun sentences

In the English-Spanish direction, a series of “*I am a + noun*” sentences was submitted for automatic translation. All sentences offered a double translation choice, with feminine first (*I am a surgeon.* > *Soy cirujana./Soy cirujano.*) unless an adjective was written before the noun, in which case the pronoun’s gender seemed arbitrary. For example, both *weak* and *strong* before *surgeon* were male.

English	Google Translate into Spanish	Gender
<i>I am a beautiful surgeon.</i>	<i>Soy una hermosa cirujana.</i>	<i>Feminine</i>
<i>I am a beautiful driver.</i>	<i>Soy una hermosa conductora.</i>	<i>Feminine</i>
<i>I am a beautiful judge.</i>	<i>Soy un juez hermoso.</i>	<i>Masculine</i>
<i>I am a beautiful president.</i>	<i>Soy un hermoso presidente.</i>	<i>Masculine</i>

TABLE 3. Gender variations of I am a beautiful + profession noun

English adjectives with gender connotations as taken from the European Institute for Gender Equality guidelines were translated into Spanish and fared better in terms of gender. Most were translated in both forms or neutrally in Spanish (*estridente*, *hormonal*), except *frigid*, which was masculine (*Soy frígido*).

Short sentences involving career nouns in the “*Es + noun*” form translated from Spanish

to English by Google Translate offered predominantly masculine translations 95% of the time while English-Spanish translations written as “*I am a + noun*” included both genders 57% of the time as opposed to 19% only female and 24% only male.

Clichéd invariable adjectives translated from Spanish to English were automatically assigned masculine gender without an option to choose from, unlike in the reverse language combination. In plural ungendered forms, Spanish translations were invariably masculine (*Young people*>*Los jóvenes*).

When the basic test sentence was followed by *is a man* or *is a woman*, the result only showed masculine gender, without stating this fact. Thus, *The inspector*>*La inspectora/El Inspector*, but *The inspector is a woman*.>*El inspector es una mujer*. Some nouns did not produce both genders either in the shorter form, like *Maoist*, *fundamentalist* or *donor*, but were always masculine. Only 11% of the sentences in this format were translated in feminine only (*activist*, *beautician*, *interpreter*, *teacher*, *survivor* and *suffragette*) and the rest in masculine gender only. By contrast, if the basic sentence is followed by *is her*, the results appear in both genders 58% of the times (*The member is her*.>*La miembro es ella./El miembro es ella.*), 25% only masculine and 16% only feminine. Nevertheless, when the list of sentences was entered together for translation, only one gender was offered and 58% was masculine.

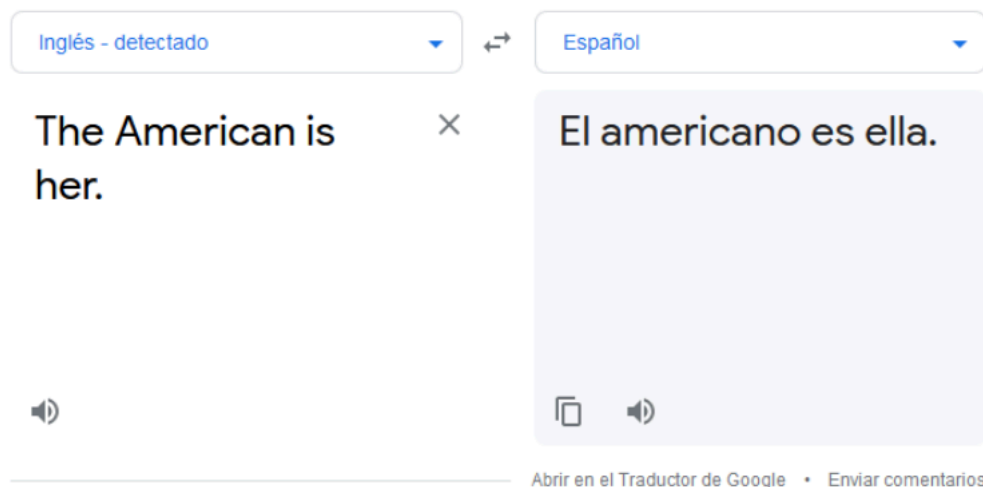


Figure 1. Print screen of Google Translate’s result for The American is her.

Some gender-specific translations appear senseless, since only one of the genders can be correct as inferred by the pronoun *his* or *her*. In the *is him* version, 64% of the results include both genders, whereas 35% are masculine and 0,02% are feminine. Therefore, the use of a

masculine pronoun produced more accurate gender results.

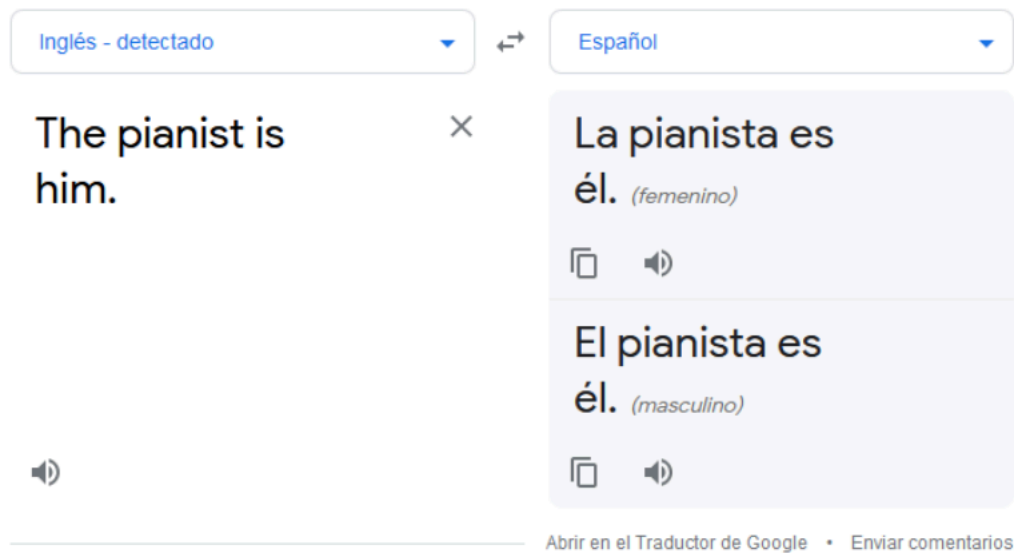
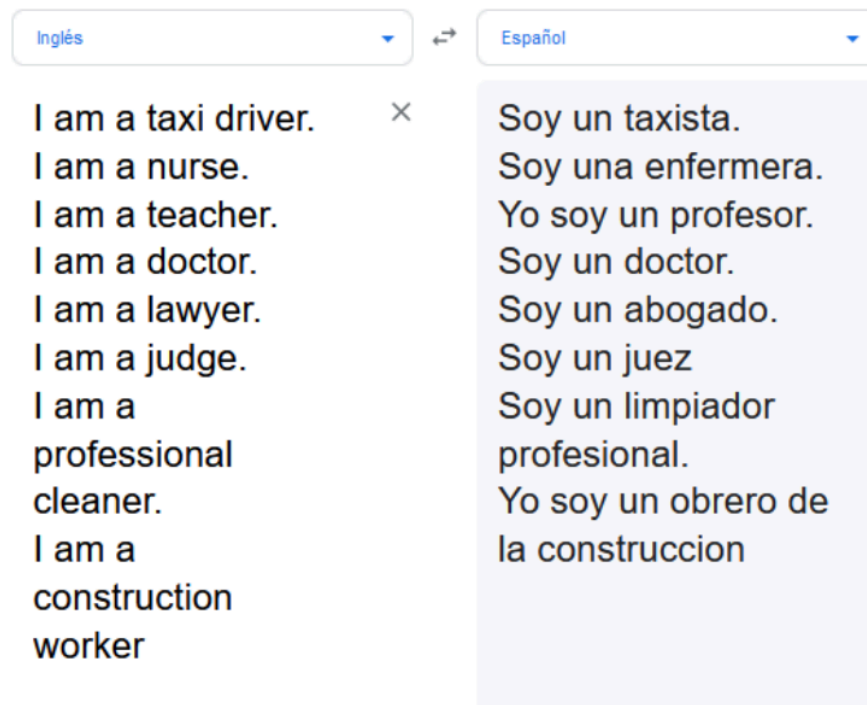


Figure 2. Print screen of Google Translate's result for The pianist is him.

Despite the recent gender-specific updates to Google Translate having showed an improvement in reducing this bias, the service is skewed towards masculine and seems to offer gender-specific translations unnecessarily when only one gender is correct, as in *The pianist is him.* > *La pianista es él./El pianista es él.*

Google Translate's English-Spanish machine bias of profession nouns does not match statistics of the gender division of jobs. In our test, *taxi driver*, *teacher*, *doctor*, *lawyer*, *judge*, *professional cleaner* and *construction worker* were all translated in masculine form (wrongly adding the article *un* which is not idiomatic in Spanish and sometimes also the unnecessary personal pronoun *yo*); only *nurse* was feminine. According to the National Statistics Office (INE), in Spain 84% of nurses, 82% of primary teachers (and 98% of infant teachers), 52% of doctors, 53% of judges and 93% of professional cleaners are women, as opposed to 44% of lawyers, 8% of construction workers and 4% of taxi drivers. Google Translate seems to apply masculine default irrespectively when on a list but the translation of individual sentences offered both female and male results, sometimes without a pronoun or article (*I am a judge.* > *Soy jueza./Soy juez.*), which is more appropriate, as in Spanish the indefinite article is not used before occupations and subject pronouns are rarely used.



Major recent studies like Prates, Avelar and Lamb’s had already shown that Google Translate underestimates the frequency of female workers and suggest debiasing algorithms that can potentially be applied at a small cost. Machine translation services need to remove masculine bias from their algorithms, especially when it comes to occupations, since non-sexist language is considered a tool for gender equality that can enable or hinder women’s access and promotion in the working environment, but offering both genders as the norm in translation is not a solution when only one can be correct. Training machine translation software to infer semantic gender from pronouns and word terminations, and applying gender-tagging in main multilingual corpora could help.

## References

- Bond, Esther. (2020) “Google fixes gender bias in Google Translate (again)”. *Slator*. <https://slator.com/machine-translation/google-fixes-gender-bias-in-google-translate-again/>
- Crespo Miguel, Mario & Marta Sánchez-Saus Laserna. (2016) “Graded acceptance in corpus-

- based English-to-Spanish machine translation evaluation. In: *CILC 2016. EPiC Series in language and linguistics*, pp. 58-70.
- European Institute for Gender Equality. (2019) “Using different adjectives for women and men”. In: EIGE. 2019. *Toolkit on gender-sensitive communication*. Luxembourg: Publications Office of the European Union, p. 24. <https://eige.europa.eu/publications/gender-sensitive-communication/challenges/stereotypes/using-different-adjectives-women-and-men>
- Google. (2021) *Google Translate help*. <https://support.google.com/translate#topic=7011755>
- Guillou et. al. (2016). “Shared task: Cross-lingual pronoun prediction”. *ACL 2016 First conference on machine translation (WMT16)*. <http://statmt.org/wmt16/pronoun-task.html>
- Johnson, Melvin. (2018) “Providing gender-specific translations in Google Translate”. *Google AI blog*. <https://ai.googleblog.com/2018/12/providing-gender-specific-translations.html>
- Johnson, Melvin. (2020) “A scalable approach to reducing gender bias in Google Translate”. *Google AI blog*. <https://ai.googleblog.com/2020/04/a-scalable-approach-to-reducing-gender.html>
- Kayser-Bril, Nicolas. (2020) “Female historians and male nurses do not exist, Google Translate tells its European users”. *Algorithm Watch*. <https://algorithmwatch.org/en/story/google-translate-gender-bias/>
- Kuczmariski, James. (2018) “Reducing gender bias in Google Translate”. *Google blog*. <https://blog.google/products/translate/reducing-gender-bias-google-translate/>
- Monti, Johanna. (2020) “Gender issues in machine translation. An unsolved problem?”. In: Von Flotow, Luise & Hala Kamal (eds.) 2020. *The Routledge handbook of translation, feminism and gender*. London: Routledge, pp. 457-468.
- O. R. Prates, Marcelo; Pedro H. Avelar & Luís C. Lamb. (2020) “Assessing gender bias in machine translation: a case study with Google Translate”. *Neural computing and applications* 32, pp. 6.363-6.381.
- Schiebinger, Londa. (2014). “Machine translation: analyzing gender”. *Gendered innovations*. <http://genderedinnovations.stanford.edu/case-studies/nlp.html#tabs-1>

Weaver, Warren. (1949) "Translation". *Machine translation of languages* 14, pp. 15-23.  
<http://www.mt-archive.info/Weaver-1949.pdf>